Estimating Numbers of Unsheltered Homeless People Through Plant-Capture and Postcount Survey Methods

Kim Hopper, PhD, Marybeth Shinn, PhD, Eugene Laska, PhD, Morris Meisner, PhD, and Joseph Wanderling, MA

Objectives. We sought to increase the accuracy of New York City's estimates of its unsheltered homeless population.

Methods. We employed 2 approaches to increasing count accuracy: a plant-capture strategy in which embedded decoys (or "plants") were used to estimate the proportion of visible homeless people missed by enumerators and a postcount survey of service users designed to estimate the proportion of unsheltered homeless people who were not visible.

Results. Plants at 17 sites (29%) reported being missed in the count, because counters either did not visit those sites or did not interview the plants. Of 293 homeless service users who were not in shelters, 31% to 41% were in locations deemed not visible to counters.

Conclusions. Both plant-capture estimation and postcount surveys are feasible approaches that can increase the accuracy of estimates of unsheltered homeless populations. (*Am J Public Health.* 2007;97:1438–1442. doi:10.2105/AJPH.2005.083600)

Communities are required to include estimates of local homeless populations in applications for US Department of Housing and Urban Development (HUD) Continuum of Care funding for programs responding to homelessness. Typically, few problems are involved in counting those individuals who are sheltered each night; methods to ensure that counts are not duplicated over time are also available. However, estimating numbers of homeless people living on the street, in parks, or in "unconventional" housing is another matter. The 2 strategies described here were used to assist New York City's ongoing efforts to improve methods of estimating the size of its unsheltered homeless population.

HUD endorses 2 methods of obtaining point-in-time counts of unsheltered homeless people: directly counting people in public places or screening those using selected services to determine whether they are homeless and without shelter.[1] Counts of visibly homeless individuals miss unsheltered people who remain out of sight during the counts.[2,3] They also depend on enumerators determining which individuals should be counted as homeless, and such judgments, whether made through observations or interviews, are subject to a host of inaccuracies. Conversely, many people using such services as soup kitchens are not homeless. As a result, surveys of their users must determine whether the arrangements of these individuals on the night in question meet operational definitions of homelessness.[4]

For purposes of its annual street count, New York's Department of Homeless Services divides the entire city, including transportation hubs and the subway system, into small zones (i.e., into a few contiguous blocks or a subway station). The department relies on service providers and police (as well as the experience gained during counts conducted in previous years) to classify each zone as high or low density, reflecting whether homeless people are likely to be found there in the middle of the night.

Volunteer counting teams are sent to all high-density zones and to a random sample of low-density zones between midnight and 4 AM on a single night in late February. Counting teams are instructed to interview all people encountered awake in their zones, to determine whether they are homeless and without shelter, and to count those found asleep. The final estimate of the number of unsheltered homeless people is the sum of those actually counted in high-density zones and a statistical extrapolation from those counted in the random sample of low-density zones.

Street counts were conducted in Manhattan and some subways in 2003, expanded to include Brooklyn and Staten Island in 2004, and extended to all 5 boroughs in 2005. As part of the ongoing efforts of the Department of Homeless Services to improve the counts, the department invited independent researchers, with federal funding from HUD, to address 2 potential limitations of the street-count methodology: first, that volunteer teams are unlikely to find and count all visibly unsheltered people on their assigned circuits and, second, that some unsheltered homeless people are likely to be in places that would be missed during the Homeless Outreach Population Estimate (HOPE) survey. We conducted 2 studies to estimate the consequences of each of these limitations for the accuracy of New York City's street count.

METHODS

Plant-Capture Technique

To address visibly unsheltered people being missed, we used a "plant-capture" variation of the more familiar capture–recapture methods employed in wildlife estimation studies, in which embedded decoys were used to estimate the proportion of visible homeless people missed by enumerators. Developed by Laska and Meisner,[5] this technique was field tested and refined as part of the unsheltered portion of the Census Bureau's 1990 "S-Night" (street and shelter) count of the homeless population of New York. Researchers placed 127 plants at 41 street and subway locations in 4 census districts; an estimated 59% were actually counted.

However, site definitions were ambiguous, enumerators missed some designated locations, and enumerators sometimes counted by observing rather than by conducting

interviews.[6–8] In this study, our use of detailed maps enabled us to estimate with increased accuracy the error introduced by the latter 2 problems. Expanded site coverage and the decision to proceed despite a weather delay enhanced the study's external validity and real-world relevance.

We deployed paid plants throughout the 5 boroughs in configurations intended to track the anticipated distribution of homeless people by borough and to conform to the city's sampling fractions of low-density sites. If the distribution of plants mimicked that of their homeless counterparts, and if plants were in the same areas to be visited by the HOPE teams, the proportion counted among this embedded population should yield a probabilistic estimate of the proportion counted among homeless people in readily accessible public spaces. Our results allowed for upward adjustment of enumerated estimates to account for people missed.

Prospective plants were recruited from universities and local service delivery staffs. Advance orientation meetings were held to review the design and conduct of the study, answer questions, and attend to paperwork. All but a few sites were field inspected during the week before the count, and detailed maps were drawn showing where plants should position themselves. Logistics were coordinated by 6 local staging centers, each staffed by 4 to 6 volunteers.

In 2005, the city postponed the count for a week, until the night of March 7–8, because of extreme weather. The ranks of hired plants were reduced to 115 as a result. Another 4 plants were recruited from our administrative ranks, enabling us to cover a total of 58 sites (with 2 plants to a site after exclusion of 1 site where plants arrived an hour late). Most plants complied with requests to dress down. Some arrived with blankets or newspapers, many accessorized, and more than a few proved adept at bringing off the disguise. Surface adjustments to appearance aside, university recruits could not hide their youth and health; these characteristics, in the eyes of HOPE's volunteer counters, may have lessened their credibility as decoys.

Postcount Interviews of Service Users

We addressed the second limitation of the street-count methodology (unsheltered homeless people being in places that would be missed) by conducting a survey (based on the methodology of the National Survey of Homeless Assistance Providers and Clients[9]) of service users on the 2 days following the count (March 8 and 9). We used published lists to construct a random sample of representatives of 5 types of services, stratified according to borough and frequency of service. We collected data at 12 of the 15 soup kitchens contacted (7 of which provided frequent services and 5 of which provided infrequent services), 2 mobile food programs, 4 of 10 drop-in centers, and all 5 of the city's largest street-outreach programs (1 was eliminated because of a misunderstanding about inclusion). The city provided data from shelter intake centers (treated as a single group). The final group aside, half of the sites were located in Manhattan and half in the outer boroughs. The total count was 23 sites.

Working rapidly, interviewers obtained informed consent from respondents (all service users or, at large sites, random samples of users) and asked them where they had spent the night of the count. Respondents were classified as not homeless, sheltered, or homeless and unsheltered during the count. The homeless and unsheltered category included people on the street, on subway trains, in subway stations or tunnels, in transportation hubs, in parks, in abandoned buildings, in indoor places not intended for sleeping, and in "other" locations.

Additional questions tailored to different locations were used to determine whether respondents could have been seen by HOPE counters (e.g., people on subways were asked whether they rode to the end of the line, where counters were stationed). Classifications were augmented by collateral reports (e.g., from outreach workers). Respondents were also asked about other services they had used in the preceding 24 hours. We intended to use these data to correct for any disproportionate sampling of services, but this proved unnecessary.

Statistical Approach

The probability of plants being counted in each stratum, $\gamma_{stratum}$, was the product of the probability that a site with plants was covered (i.e., visited) by enumerators (p_{cov}) multiplied by the probability of plants being captured (i.e., counted), given that the site was covered (p_{cap}), multiplied in turn by the sampling fraction used by the city in selecting sites (p_{sample}). We estimated the first 2 probabilities by using the reported experiences of plants on the night of the count (whether a given site was visited and, if so, whether plants were counted). We estimated the total street count by summing estimates for high-density and low-density strata. As a result of the small number of subway sites, the p_{cov} value for high-density surfaces and subways was a weighted average. The formula for the plant-capture adjusted estimate was as follows:

$$(1) \qquad N_{est} = \sum \frac{\text{HOPE count in stratum}}{\gamma \text{ stratum}},$$

where the sum was taken over all of the strata.

Similarly, the postcount survey allowed for estimates of the probability of individuals being visible and thus being counted ($p_{visible}$). Three strata (subways, Manhattan, and an aggregate of the other boroughs) were constructed, reflecting similarities in reported visibility within and differences between those strata. We were able to calculate an overall population estimate by aggregating the corresponding plant-capture adjusted estimates from HOPE into these groupings, dividing by the estimated $p_{visible}$ value for each stratum, and then summing over the different strata. We used a Bonferroni correction for the 9 estimated parameters (strata × density × visibility) to calculate 95% confidence intervals.

RESULTS

Estimating Numbers of Visibly Unsheltered Individuals

Plants in 17 of the 58 eligible sites were neither interviewed nor, to the best of their knowledge, counted. (If interviewed, plants handed over a site-specific sticker to the interviewer.) Both the uncounted and the counted categories included a variety of close calls and ambiguous misses; among counters, for example, rules for approaching people were applied irregularly. In addition, boundary ambiguities plagued even well-defined sites (such as parks).

TABLE 1—Estimated Probabilities From Plant-Capture Count and Postcount Survey: New York City, 2005

	Number of Sites With Plants		Estimated Probability of Site Being Covered		Conditional Probability of Plants Being Counted		Fraction of Sites Sampled by City		Estimated Probability of Service User Definitely Being Visible (95% CI)	Estimated Probability of Service User Possibly Being Visible (95% CI)
	HD	LD	HD	LD	HD	LD	HD	LD		
Manhattan	24	2	0.75	1.00	1.00	1.00	241 of 241	91 of 701	0.58 (0.49, 0.67)	0.68 (0.59, 0.76)
Outer boroughs	20	2	0.95	1.00	0.79	1.00	268 of 268	311 of 5597	0.33 (0.25, 0.43)	0.51 (0.41, 0.60)
Subways	10	0[a]	0.80		0.63		98 of 98	47 of 301	0.93 (0.85, 0.97)	0.94 (0.86, 0.97)

Note. HD = high density (based on expectations of finding at least 2 [Manhattan] or 1 [rest of New York City] homeless persons in a given zone); LD = low density (no homeless persons expected to be found in a given zone); CI = confidence interval.
[a]Only a single low-density subway site was included in our plant-capture sample, so these data were pooled with the data from the high-density subway sites.

Twice, counters gave plausible accounts of why plants they did not interview might not have been homeless (a flirtation in progress and an undercover stakeout). In several instances, plants proved overly helpful, beckoning to counters who would otherwise have passed them by. Especially difficult and extensive were ingrained dispositions that led counters to not see certain people as homeless, a phenomenon we labeled "discounting." The strata-specific distribution of plants, the probability that a given site was covered, and the probability that plants were "captured" if the site was covered are shown in Table 1.

TABLE 2—Unsheltered Service Users' Visibility Status, by Location: New York City, 2005

	Visibility Status			
	Not Visible, No. (%)	Visible, No. (%)	Uncertain, No. (%)	Total, No.
Street	22 (22)	70 (69)	10 (10)	102
Subway train	1 (1)	67 (97)	1 (1)	69
Subway tunnel or station	4 (31)	9 (69)	0 (0)	13
Transportation hub	0 (0)	15 (100)	0 (0)	15
Park	5 (50)	4 (40)	1 (10)	10
Abandoned building	14 (100)	0 (0)	0 (0)	14
Indoor place	28 (100)	0 (0)	0 (0)	28
Other (e.g., porch, car, ferry)	16 (38)	9 (21)	17 (40)	42
Total	90 (31)	174 (59)	29 (10)	293

Estimating Visibility Status Among Service Users

We interviewed 1171 respondents at 23 service sites; an additional 378 respondents were not interviewed because they refused, did not speak English, or were asleep. A total of 314 (27%) respondents reported being homeless and unsheltered during the count. That percentage varied markedly (and unsurprisingly) according to type of service. Using location descriptions provided by 293 unsheltered service users with usable data, amended in some cases by collateral reports, we classified 174 (59%) of these individuals as definitely visible, 90 (31%) as definitely not visible, and 29 (10%) as uncertain in terms of visibility status. Locations not visible included indoor places such as stairwells and boiler rooms, all-night commercial establishments, abandoned buildings, and hidden outdoor locations (e.g., porches hidden by shrubbery). Locations labeled "uncertain" included vehicles and makeshift shelters, among others (Table 2).

To determine whether respondents' visibility varied according to surface (subway vs street), borough (Manhattan vs outer boroughs), or type of service at which they were sampled, we estimated a series of 2-level hierarchical linear (random regression) models to account for sample clustering.[10] Surprisingly, although type of service at which individuals were sampled was highly related to whether they were homeless and unsheltered, it was unrelated to that group's likelihood of being visible to enumerators. (Thus, it was not necessary to apply differential weighting by type of service to obtain an overall estimate of population size.) Two factors were associated with visibility: respondents in Manhattan were more likely to have been visible during the count, and those on the street during the count were substantially less likely to have been visible than those in the subway system.

On the basis of the postcount survey, the last 2 columns of Table 1 present estimates (and their 95% confidence intervals) of the probability that unsheltered homeless service users were definitely visible during the count and of the probability that they at least might have been visible.

DISCUSSION

We used 2 different strategies designed to improve the accuracy of street counts of homeless people in New York City. The plant-capture method provided a valid method of making statistical adjustments to counts; the undercounts it revealed could in principle be corrected by improving procedural rigor. By contrast, our survey of service users suggesteds that counting only individuals visible on the street or in subways cannot, in itself, provide an accurate estimate of the size of the homeless population living on the street, in parks, or in places not intended for habitation.

Given that, as described earlier, an unknown number of missed plants could have been "discounted" because they in fact did not appear to be homeless, the city argued that our original count should be adjusted by

between 15% (combined p_{cov}) and 29% (combined p_{cap}). The city took the midpoint and adjusted by 22% for an official estimate of 4395; ours was 4630. We considered the postcount survey as producing preliminary findings, and the data derived from this survey were not used to adjust the estimate. Such an adjustment would have required additional information on patterns of service usage, and the further assumption that, among unsheltered homeless people, service users are at least as likely as nonusers to be visible during the count. Assuming our sample was representative, adjustments based on both missed plants and lack of visibility among service users may be made to the count to obtain a more accurated number of unsheltered homeless people.

The plant-capture strategy yielded other findings as well. Informal reports suggested that the use of this technique had felicitous effects; for example, the mere existence of a "quality control" study probably improved the diligence of working volunteers, at least those from training sites where the study was acknowledged and stressed. As a result of the ease with which plants could be placed in the territory to be canvassed and readily unmasked if discovered, efficiency costs were negligible.

More striking was the apparent ease with which some of the counting teams deviated from their instructions. In particular, reports of counters walking entirely past plants typify the practice of discounting, that is, ignoring certain street occupants because "they don't look homeless" and instead, according to some tacit guide, approaching only those who do. However, discounting is somewhat tricky: in a few instances, although plants were ignored, other site occupants were not; as noted earlier, counters gave plausible (if mistaken) readings of why plants should not be approached; and, in the subways, the logistics of approach made it prudent to engage in triage.

We cannot state with certainty the frequency at which actual homeless people who departed from stereotypes were missed, but the survey of service users supplies some evidence. We did not ask homeless service users whether they had been counted; those missed could have been in low-density areas that were not sampled. However, 1 homeless woman reported that counters interviewed others but not her, because they did not think she was homeless. A second respondent reported spending the night quite visibly in an airport terminal, but the team canvassing the airport reported no one. Both reports suggest that homeless respondents, similar to plants, were sometimes discounted.

With respect to the service-user survey, the most striking finding is that a substantial proportion of homeless individuals reported being in places during the HOPE street count where they could not have been seen by the volunteer enumerators. This was especially true of surface sites in the outer boroughs, which provide more hidden places to stay (e.g., abandoned buildings, vehicles, porches, backs of buildings) than are available in Manhattan, where most buildings are flush to sidewalks and fences often prevent access to backs of buildings.

Service users in the subway system were most likely to be visible. Our results validate the city's strategy of counting homeless subway riders at terminal stations, because almost everyone on the trains traveled to the end of the line. (However, according to our plant-capture findings, the probability of individuals in fact being counted when in that system was not high, but our numbers were small.) Because relatively few other cities in the country have all-night subway systems, and because most streetscapes resemble New York's outer boroughs more than Manhattan, the data on differences in visibility suggest that street estimates may be substantially more inclusive in New York than in other cities. If we take the outer borough surface figure as representative of other cities, it could be that even well-executed street counts elsewhere miss half or more of homeless people.

As with all surveys, ours relied on self-reported data (modified in a few cases by information from outreach workers). We suspect that the accuracy levels associated with classifying individuals as either homeless or not homeless are lower than those associated with judging the reported visibility of unsheltered homeless individuals. Once one has reported the potentially embarrassing fact of homelessness, reporting where one spent the previous night seems less fraught with implications for self-esteem. We only asked that respondents provide the types of locations they frequented; we did not ask them to report specific hiding places.

A more serious limitation is that a sample of service users may not accurately represent the overall population of unsheltered homeless people. However, if service users are at least as likely to be visible as nonusers, an estimate based on service users provides a lower bound of the true size of the population. Also, without information on overall patterns of service usage by the unsheltered homeless, it is not clear how to use the service-specific visibility rates reported by a sample of users to adjust the population-size estimate. Although, in our study, visibility did not depend on type of service used, this finding should be replicated before assuming its applicability to other counts, even in New York.

Our findings suggest that even such aggressively mounted street counts as HOPE are fated to miss a large portion of unsheltered homeless individuals for a pair of reasons. First, as a result of operational irregularities, coupled with such contingencies as weather and logistical limitations, coverage will always fall short of complete. An embedded plant-capture strategy can help provide a more valid estimate. In our study, the yield of this method was compromised by reduced numbers of plants, uneven training of volunteer counters, and the apparently widespread practice of discounting. Plant recruiting can be intensified and training improved. Plants who "play to" stereotypes (for example, with tattered outerwear, blankets, and feet wrapped in plastic bags) may be less likely to be discounted. Informing counters that they should seek out plants who may not "appear homeless" may reduce discounting of genuine homeless people who would otherwise pass as housed.

The second, less easily remedied source of error resides at the heart of street survival strategies. To the degree that invisibility serves the purpose of security and uninterrupted sleep, substantial proportions of unsheltered homeless individuals are likely to favor it. Even easily accessed but visibility-blocking structures (such as shanty settlements) may be excluded from counts in the interest of volunteer safety. (For the same

reason, poorly lit parks will be only sketchily explored.) Next-day surveys of service users can further increase the validity of estimates and help officials gauge the extent of such out-of-sample populations, at least to the degree that the "invisibly" homeless routinely use available services. Other strategies sometimes recommended, such as relying on outreach workers for estimates, are unlikely to be effective if, as was true here, only a minority of service users report contact with such workers.

It is important to recognize that 1-night street counts represent only a small proportion of those affected by homelessness. The average daily census in New York shelters in 2005 (including families) was 33 687,[11] or 0.42% of the city's population. In addition, an unduplicated count based on shelter records showed that 3.27% of the city's population used shelters over the 5-year period from 1988 through 1992,[12] when the nightly shelter count averaged only about 70% of the count for 2005. Although it may be a considerable challenge, obtaining period prevalence estimates of numbers of unsheltered homeless individuals, such as yearly estimates, may be more important than improving point prevalence estimates on a given night. Yearly estimates would also be less subject to vagaries of weather and would provide a more reliable base for judging the success of efforts to reduce street homelessness. ■

About the Authors

Kim Hopper is with the Nathan Kline Institute for Psychiatric Research, Orangeburg, NY, and the Mailman School of Public Health, Columbia University, New York City. Marybeth Shinn is with the Steinhardt School of Culture, Education and Human Development, and the Wagner School of Public Service, New York University, New York City. Eugene Laska and Morris Meisner are with the Nathan Kline Institute for Psychiatric Research, Orangeburg, and the Department of Psychiatry, New York University Medical School, New York City. Joseph Wanderling is with the Nathan Kline Institute for Psychiatric Research, Orangeburg.

Requests for reprints should be sent to Kim Hopper, PhD, Nathan Kline Institute for Psychiatric Research, 140 Old Orangeburg Rd, Orangeburg, NY 10962 (e-mail: hopper@nki.rfmh.org).

This article was accepted January 7, 2007.

Contributors

K. Hopper, E. Laska, M. Meisner, and J. Wanderling designed the plant-capture study; M. Shinn designed the postcount study. K. Hopper and M. Shinn managed the field implementation of both studies. All of the authors participated in the analyses.

Acknowledgments

For indispensable collaboration that made this work possible, we thank our colleagues in the New York City Department of Homeless Services, especially Maryanne Schretzman, Jay Bainbridge, Katie Appel, and Şerban Iorga. For operational assistance on the count and survey nights, we thank our colleagues in the Statistics and Service Research Department of the Nathan Kline Institute for Psychiatric Research. For extraordinary logistical proficiency in recruitment and organization, we thank Bahar Kumar.

Human Participant Protection

No protocol approval was needed for the plant-capture study; New York University approved the protocol for the postcount interviews.

References

1. US Dept of Housing and Urban Development. A guide to counting unsheltered homeless people. Available at: http://www.hud.gov/offices/cpd/homeless/library/countinghomeless/countingguide.pdf. Accessed July 27, 2007.

2. Link B, Phelan J, Bresnahan M, Stueve A, Moore R, Susser E. Lifetime and five-year prevalence of homelessness in the United States: new evidence on an old debate. *Am J Orthopsychiatry.* 1995;65:347–354.

3. Wright JD, Devine JA. Housing dynamics of the homeless: implications for a count. *Am J Orthopsychiatry.* 1995;65:320–329.

4. Campanelli P, Salo M, Schwede L, Martin E. *Research on Enumerating Homeless Persons: Results of a Census Bureau Test of Alternative Methods.* Washington, DC: Center for Survey Methods Research, US Census Bureau; 1990.

5. Laska EM, Meisner M. A plant-capture method for estimating the size of a population from a single sample. *Biometrics.* 1993;49:209–220.

6. Hopper K. Counting the homeless in New York: an ethnographic perspective. *N Engl J Public Policy.* 1992;8:771–791.

7. Hopper K. Counting the homeless: S-night in New York. *Eval Rev.* 1992;16:376–388.

8. Martin E, Laska E, Hopper K, Meisner M, Wanderling J. Issues in the use of a plant-capture method for estimating the size of the street dwelling population. *J Off Stat.* 1997;13:59–73.

9. Burt M, Aron LY, Douglas T, Valente J, Lee E, Iwen B. *Homelessness: Programs and the People They Serve: Findings of the National Survey of Homeless Assistance Providers and Clients.* Washington, DC: Urban Institute; 1999.

10. Raudenbush S, Bryk T, Cheong YF, Congdon R. *HLM 5: Hierarchical Linear and Nonlinear Modeling.* Lincolnwood, Ill: Scientific Software International Inc; 2000.

11. New York City Government. Historic data report. Available at: http://www.nyc.gov/html/dhs/downloads/pdf/histdata.pdf. Accessed July 27, 2007.

12. Culhane DP, Dejowski EF, Ibaez J, Needham E, Macchia I. Public shelter admission rates in Philadelphia and New York City: the implications of turnover for sheltered population counts. *Housing Policy Debate.* 1994;5:107–140.